USING STRUCTURAL INFORMATION FOR PREDICTING NAD(P)(H) COFACTOR SPECIFICITY, WHILE UNVEILING THE RESPONSIBLE MOLECULAR DETERMINANTS, IN ENZYMES WITH UNKNOWN STRUCTURE

Tiago Resende, Centro de Engenharia Biológica, Universidade do Minho
tiagoresende@ceb.uminho.pt
Isabel Rocha, Centro de Engenharia Biológica, Universidade do Minho
Cláudio Soares, ITQB NOVA, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa

Key Words: Enzyme structural engineering, cofactor engineering

Systems biology foundations broadly rely upon well performed gene homology annotations. As metabolic model reconstructions become a relevant tool for performing fundamental studies and bioprocess design, the impact of accurate enzymatic function assignments becomes evident. Uncertainty on the usage of NADP(H) or NAD(H) as co-factors has a major impact in metabolic engineering applications, severely affecting both predictions and strain design results.

In this work, we unveil the molecular determinants for cofactor specificity, using enzyme structural information. To do so, we have created a representative dataset of enzymes present in Protein DataBase with NAD(P)(H) as cofactors and retrieved information on every interaction between aminoacid residues and cofactor atoms, storing numerous data on binding site properties. The ensuing data analysis culminated on the identification of aminoacid residues with a significantly higher amount of binding contacts with specific cofactor atoms.
These findings where successfully applied to enzymes not structurally characterized, by using protein modelling and machine learning algorithms. To enable high throughput cofactor preference prediction, a method was developed to automatically attribute cofactor specificity preference, when given the aminoacid sequence only, whilst returning the modelled structure of the query sequence.

We believe these results represent an important contribution for metabolic engineering, with the enhancement of the sensitivity and reliability of metabolic models, through the lower input of erroneous or redundant reactions, improving the overall performance of metabolic simulations.